In order to fully grasp the importance of a study’s findings and limitations, it is important to understand some key concepts related to statistical significance, power, and error. In this article, we will discuss how to interpret research findings by discussing p-values and statistical significance. We will then discuss common types of errors encountered in pharmacy research (Type I and Type II errors), and the importance of statistical power to research. Finally, we briefly discuss the difference between statistical and clinical significance.

Probability and Chance
Biomedical research usually starts with a question such as “How does Drug X affect A1c in patients with Type II diabetes?” Next, researchers identify the population of interest and select an appropriate sample from that population. A population (often...
Probability is an estimation of how likely it is that something will happen or that a statement is true. The probability of a particular event occurring ranges from 0 (never occurs) to 1 (always occurs). Suppose, for example, we wanted to predict the outcome of flipping a coin. Since a coin has two sides, we know there are only two possible outcomes: heads or tails. Therefore, the probability of the coin landing heads up is 1/2 (or 0.5), and the probability of the coin landing tails up is 1/2 (or 0.5). Similarly, if we were to roll a 6-sided die, the probability of rolling a 2 is 1/6, the probability of rolling an even number is 3/6 (or 1/2), and so on.

Now suppose we flipped a coin 50 times and got 50 heads in a row. Since we know the probability of getting a heads is 1/2, the probability of getting 50 consecutive heads is very low. According to this low probability value, most reasonable people would conclude that heads were favored; that is, that getting a heads was more likely than getting a tails. This illustrates a very important principle: that if, under a given assumption (e.g., the probability of a heads is 1/2), the probability of a particular sample's results is exceptionally small (e.g., 50 straight heads), one would conclude that the assumption is incorrect.

### Hypothesis Testing and P-Values
A hypothesis is a claim or statement about a property of a population. For example, “The manufacturer of Drug X claims it reduces A1c by 1.0% in patients with diabetes.” Research involves two types of hypotheses: a null hypothesis and an alternative hypothesis. In biomedical research, the null hypothesis, typically denoted \( H_0 \), is the claim that the intervention has no effect. The alternative hypothesis, typically denoted \( H_1 \), is the claim that the intervention actually does have an effect. In this case: \( H_0 \): Drug X has no effect on A1c. \( H_1 \): Drug X has an effect on A1c.

To investigate a hypothesis, we analyze a sample through a clinical study to determine if the results are reasonably consistent or conflicting with the hypothesis. This is commonly done using a p-value. The p-value is the probability of obtaining results as extreme or more extreme than the results observed given \( H_0 \) is true. For example, assuming our \( H_0 \) is true that there is no effect of the study medication, the p-value is the probability of obtaining the observed (or larger) difference in A1c in our intervention group compared to the control group. (This can also be done using confidence intervals, which was discussed in part 3 of this series. The statistical tests used to determine the p-value will be covered in part 5 of this series.) Because there is inherent variability or error in the way things are measured, we define a pre-specified “acceptable” error level, typically designated using \( \alpha \). The biomedical standard for an acceptable error level is 5% (or \( \alpha = 0.05 \)), although in some instances a lower (1% or \( \alpha = 0.01 \)) or higher (10% or \( \alpha = 0.1 \)) level of error may be acceptable. An \( \alpha \) level of 0.05 means there is a less than 1:20 (5%) chance that the results occurred due to random error.

Table 1: Error Summary

<table>
<thead>
<tr>
<th>( H_0 ) True (Truly no treatment effect)</th>
<th>( H_1 ) True (Truly a treatment effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject ( H_0 ) (Treatment effect found)</td>
<td>Correct (Find a treatment effect when there is truly no effect)</td>
</tr>
<tr>
<td>Type I Error (Find a treatment effect when there is truly no effect)</td>
<td>Probability = ( \alpha ) (significance level)</td>
</tr>
<tr>
<td>Correct (Do not find a treatment effect when there is truly no effect)</td>
<td>Probability = 1 - ( \alpha )</td>
</tr>
<tr>
<td>Type II (Do not find a treatment effect when there is truly a treatment effect)</td>
<td>Probability = ( \beta ) (1 - power)</td>
</tr>
</tbody>
</table>

Despite our best efforts, errors in statistical measurement can occur. Errors are typically classified as either a Type I error or a Type II error. Errors in statistical measurement can occur. Errors are typically classified as either a Type I error or a Type II error.
II error. A Type I error occurs when $H_0$ is rejected when it is true (a “false positive”). This occurs when a researcher claims a finding was statistically significant when in fact it was not. This commonly occurs when small sample sizes are used that allow random variation to have a large effect. This can also occur when a large number of statistical tests are performed (called “repeated testing”); positive results can be found simply by undertaking so many comparisons that significant results will eventually be found by chance. A Type II error occurs when we fail to reject $H_0$ when it is false (a “false negative”). This occurs when a researcher claims there was no significant difference when there actually was. This commonly occurs when the sample size is too small for a difference to be detected or when our $\alpha$ is set too small.

Conventionally, we allow a 20% risk of a Type II error (typically designated using $\beta=0.2$); however, the actual value of $\beta$ varies with the size of $\alpha$, the size of the effect, the size of the sample, and the variance of the original distribution. Statistical power is the ability of a study to detect a statistically significant difference, and is defined as $1-\beta$. Since we typically set $\beta$ to be 20%, the statistical power is typically set at 80% ($1 - 0.2 = 0.8$), although it can be set higher. Many studies do not have adequate sample sizes to be definitive in their conclusions and are underpowered; there is actually a significant difference but the sample size is not large enough to detect it. Table 1. summarizes the different types of error.

### Statistical versus Clinical Significance

A final important concept is the difference between statistical significance and clinical significance. A statistically significant finding may not always be clinically meaningful in practice. For example, a study may find that a new diabetes drug significantly reduces patient A1c by 0.1% compared to placebo ($p=0.05$). Although this finding may be statistically significant, it may not be a clinically important finding for many patients. However, the clinical significance of a finding varies on the finding and the situation. Although for many patients a 0.1% reduction in A1c may not be clinically useful, it may be useful for a patient who has contraindications or adverse reactions to other diabetes medications.

### Summary

This article reviewed several key concepts related to statistical significance, power, and error. This information is useful to understand the importance of a study’s findings and limitations. The next article in this series will cover the use of statistical tests to determine a p-value.

---

**Practice questions**

1. Which of these statements is the correct interpretation of a p-value? An exceptionally small p-value indicates the difference between two groups is:
   a. likely due to chance
   b. unlikely to be due to chance
   c. is clinically significant
   d. is not clinically significant

2. A recent study found that Drug X reduces cardiovascular mortality by 5% compared to placebo ($p=0.08$). Assuming $\alpha=0.05$ and the following hypotheses, what conclusion would we make?
   $H_0$: Drug X is not significantly different from placebo
   $H_1$: Drug X is significantly different from placebo
   a. We reject $H_0$ and accept $H_1$
   b. We reject $H_0$ and fail to reject $H_1$
   c. We accept $H_0$
   d. We fail to reject $H_0$

3. A study of a new medication for diabetes finds a 0.8% lowered A1c in the treatment group compared to the placebo group with a p-value of 0.057. Which of the following statements regarding error is correct?
   a. Type I error could be possible
   b. Type II error could be possible
   c. Type I and II errors could be possible
   d. There is unlikely to be an error

**Answers:**

1. b. The p-value is the probability of obtaining the observed (or larger) difference in A1c in our intervention group compared to the control group. When the p-value is very small we can reject the null hypothesis that there is no difference between the two groups.

2. d. Since our p-value is greater than $\alpha$, we would fail to reject $H_0$ (no effect). We cannot accept $H_0$, only fail to reject it.

3. b. Using the conventional cutoff of 0.05 for the p-value, we find the results of this study are not statistically significant. A Type II error is possible in that there may truly be a difference but the study was underpowered to detect a difference.

### References and suggestions for further review: